Coronavirus Main Proteinase ($3CL^{pro}$) Structure: Basis for Design of Anti-SARS Drugs

Kanchan Anand *et al.*

Science **300**, 1763 (2003);
DOI: 10.1126/science.1085658

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

31. R. J. Krauzlis, C. S. Stone, *Trends Neurosci*. **22**, 544 (1999).
32. C. L. Colby, J. R. Duhamel, M. E. Goldberg, *Cereb. Cortex* **5**, 470 (1995).
33. K. Nakamura, C. L. Colby, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4026 (2002).
34. S. Ben Hamed, J. R. Duhamel, *Exp. Brain Res*. **142**, 512 (1998).
35. B. Gaymard, C. J. Ploner, S. Rivaud, A. I. Vermersch, C. Pierrot-Deseilligny, *Exp. Brain Res.* **123**, 159 (1998).
36. D. C. Somers, S. B. Nelson, M. Sur, *J. Neurosci.* **15**, 5448 (1995).
37. V. Dragoi, J. Sharma, E. K. Miller, M. Sur, *Nature Neurosci.* **5**, 883 (2002).
38. We thank P. Sinha for insightful discussions and comments and J. Schummers, J. Mariño, B. Scott, and D. Lyon for careful reading of the manuscript and suggestions. Supported by a fellowship from the McDonnell-Pew Foundation (V.D.) and grants from NIH (M.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/300/5626/1758/DC1
Materials and Methods
SOM Text
Fig. S1
References

19 December 2002; accepted 8 May 2003

Coronavirus Main Proteinase ($3CL^{pro}$) Structure: Basis for Design of Anti-SARS Drugs

Kanchan Anand,[1,2]† John Ziebuhr,[3]† Parvesh Wadhwani,[4] Jeroen R. Mesters,[1,2] Rolf Hilgenfeld[1,2]*

A novel coronavirus has been identified as the causative agent of severe acute respiratory syndrome (SARS). The viral main proteinase (M^{pro}, also called $3CL^{pro}$), which controls the activities of the coronavirus replication complex, is an attractive target for therapy. We determined crystal structures for human coronavirus (strain 229E) M^{pro} and for an inhibitor complex of porcine coronavirus [transmissible gastroenteritis virus (TGEV)] M^{pro}, and we constructed a homology model for SARS coronavirus (SARS-CoV) M^{pro}. The structures reveal a remarkable degree of conservation of the substrate-binding sites, which is further supported by recombinant SARS-CoV M^{pro}-mediated cleavage of a TGEV M^{pro} substrate. Molecular modeling suggests that available rhinovirus $3C^{pro}$ inhibitors may be modified to make them useful for treating SARS.

Human coronaviruses (HCoVs) are major causes of upper respiratory tract illness in humans; in particular, the common cold (*1*). To date, only the 229E strain of HCoV has been characterized in detail, because it used to be the only isolate that grows efficiently in cell culture. It has recently been shown that a novel HCoV causes severe acute respiratory syndrome (SARS), a disease that is rapidly spreading from its likely origin in southern China to several countries in other parts of the world (*2*, *3*). SARS is characterized by high fever, malaise, rigor, headache, and nonproductive cough or dyspnea and may progress to generalized interstitial infiltrates in the lung, requiring intubation and mechanical ventilation (*4*). The fatality rate among people with illness meeting the current definition of SARS is presently around 15% [calculated as deaths/(deaths + surviving patients)]. Epidemiological evidence suggests that the transmission of this newly emerging pathogen occurs mainly by face-to-face contact, although other routes of transmission cannot be fully excluded. By 9 May 2003, more than 7000 cases of SARS had been diagnosed worldwide, with the numbers still rapidly increasing. At present, no efficacious therapy is available.

Coronaviruses are positive-stranded RNA viruses featuring the largest viral RNA genomes known to date (27 to 31 kb). The gene for the human coronavirus 229E replicase, encompassing more than 20,000 nucleotides, encodes two overlapping polyproteins [pp1a (replicase 1a, ~450 kD) and pp1ab (replicase 1ab, ~750 kD) (*5*)] that mediate all the functions required for viral replication and transcription (*6*). Expression of the C-proximal portion of pp1ab requires (–1) ribosomal frameshifting (*5*). The functional polypeptides are released from the polyproteins by extensive proteolytic processing. This is primarily achieved by the 33.1-kD HCoV 229E main proteinase (M^{pro}) (*7*), which is frequently also called 3C-like proteinase ($3CL^{pro}$) to indicate a similarity of its cleavage-site specificity to that observed for picornavirus 3C proteinases [$3C^{pro}$ (table S1)], although we have recently shown that the structural similarities between the two families of proteinases are limited (*8*). The M^{pro} ($3CL^{pro}$) cleaves the polyprotein at no less than 11 conserved sites involving Leu-Gln↓(Ser,Ala,Gly) sequences (the cleavage site is indicated by ↓), a process initiated by the enzyme's own autolytic cleavage from pp1a and pp1ab (*9*, *10*). This cleavage pattern appears to be conserved in the M^{pro} from SARS coronavirus (SARS-CoV), as we deduced from the genomic sequence published recently (*11*, *12*) and prove experimentally here for one cleavage site (see below). The SARS-CoV polyproteins have three noncanonical M^{pro} cleavage sites with Phe, Met, or Val in the P2 position, but the same cleavage sites are unusual in other coronaviruses as well. The functional importance of M^{pro} in the viral life cycle makes this proteinase an attractive target for the development of drugs directed against SARS and other coronavirus infections.

Here we report three three-dimensional (3D) structures of coronavirus M^{pro}s, which together form a solid basis for inhibitor design: (i) the crystal structure, at 2.54 Å resolution, of the free enzyme of human coronavirus (strain 229E) M^{pro}; (ii) a homology model of SARS-CoV M^{pro}, based on the crystal structure of HCoV 229E M^{pro} described here and on that of the homologous enzyme of the related porcine transmissible gastroenteritis (corona)virus (TGEV), which we determined previously (*8*); and (iii) the 2.37 Å crystal structure of a complex between TGEV M^{pro} and a substrate-analog hexapeptidyl chloromethyl ketone (CMK) inhibitor. Comparison of the structures shows that the substrate-binding sites are well conserved among coronavirus main proteinases. This is supported by our experimental finding that recombinant SARS-CoV M^{pro} cleaves a peptide corresponding to the N-terminal autocleavage site of TGEV M^{pro}. Further, we find the binding mode of the hexapeptidyl inhibitor to be similar to that seen in the distantly related human rhinovirus 3C proteinase ($3C^{pro}$) (*13*). On the basis of the combined structural information, a prototype inhibitor is proposed that should block M^{pro}s and thus be a suitable drug for targeting coronavirus infections, including SARS.

The 2.54 Å crystal structure of HCoV 229E M^{pro} (*14*) shows that the molecule comprises three domains (Fig. 1A). Domains I and II (residues 8 to 99 and 100 to 183, respectively) are six-stranded antiparallel β barrels and together resemble the architecture of chymotrypsin and of picornavirus 3C proteinases. The substrate-binding site is located

[1]Institute of Biochemistry, University of Lübeck, D-23538 Lübeck, Germany. [2]Institute of Molecular Biotechnology, D-07745 Jena, Germany. [3]Institute of Virology and Immunology, University of Würzburg, D-97078 Würzburg, Germany. [4]Institute of Molecular Biology, University of Jena, D-07745 Jena, Germany.

*To whom correspondence should be addressed at the Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany. E-mail: hilgenfeld@biochem.uni-luebeck.de
†These authors contributed equally to this work.

in a cleft between these two domains. A long loop (residues 184 to 199) connects domain II to the C-terminal domain (domain III, residues 200 to 300). This latter domain, a globular cluster of five helices, has been implicated in the proteolytic activity of M^{pro} (*15*). The HCoV 229E M^{pro} structure is very similar to that of TGEV M^{pro} (*8*). The root mean square (rms) deviation between the two structures is ~1.5 Å for all 300 Cα positions of the molecule (*16*), but the isolated domains exhibit rms deviations of only ~0.8 Å. HCoV 229E and TGEV are both group I coronaviruses (*17*), and their main proteinases share 61% sequence identity.

For comparison of its enzymatic properties with those of the HCoV and TGEV M^{pro}s, we expressed SARS-CoV (strain TOR2) M^{pro} in *Escherichia coli* (*18*) and preliminarily characterized the proteinase. The amino acid sequence of SARS-CoV M^{pro} displays 40 and 44% sequence identity to HCoV 229E M^{pro} and TGEV M^{pro}, respectively (see Fig. 1B for a structure-based alignment). Identity levels are 50 and 49%, respectively, between SARS-CoV M^{pro} and the corresponding proteinases from the group II coronaviruses: mouse hepatitis virus (MHV) and bovine coronavirus (BCoV). Finally, SARS-CoV M^{pro} shares 39% sequence identity with avian infectious bronchitis virus (IBV) M^{pro}, the only group III coronavirus for which a main proteinase sequence is available. These data are in agreement with the conclusion deducible from the sequence of the whole SARS-CoV genome (*11*, *12*) that the new virus is most similar to group II coronaviruses, although some common features with IBV (group III) can also be detected. Others have defined SARS-CoV as the first member of a new group IV (*11*).

The level of similarity among SARS-CoV, HCoV 229E, and TGEV M^{pro}s, allowed us to construct a reliable 3D model of SARS-CoV M^{pro} (Fig. 1C). There are three one- or two-residue insertions in SARS-CoV M^{pro}, relative to the structural templates; as is to be expected, these are all located in loops and do not present a problem in model building. Interestingly, domains I and II show a higher degree of sequence conservation (42 to 48% identity) than does domain III (36 to 40%) between SARS-CoV M^{pro} and the coronavirus group I enzymes.

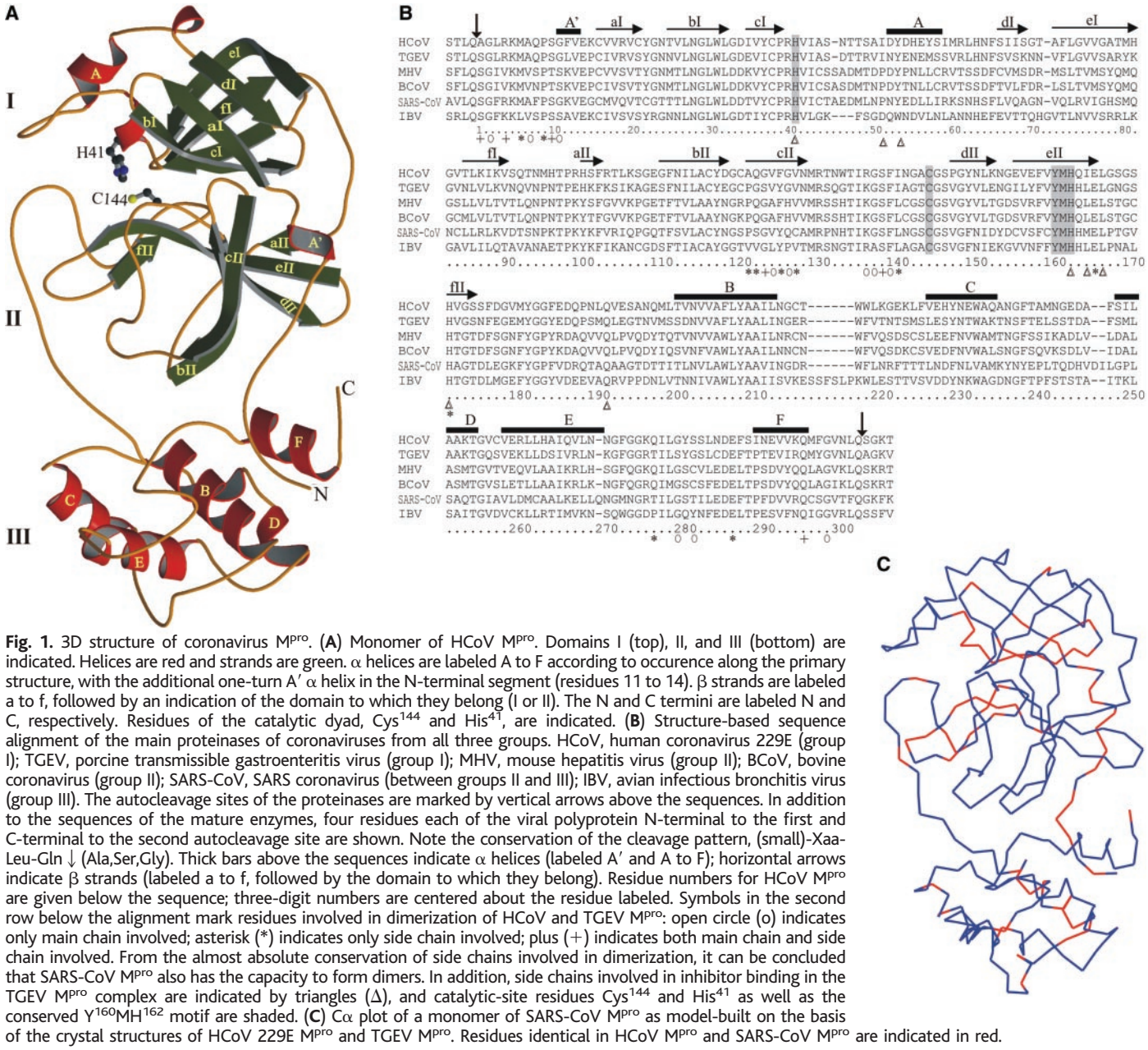


Fig. 1. 3D structure of coronavirus M^{pro}. (**A**) Monomer of HCoV M^{pro}. Domains I (top), II, and III (bottom) are indicated. Helices are red and strands are green. α helices are labeled A to F according to occurence along the primary structure, with the additional one-turn A′ α helix in the N-terminal segment (residues 11 to 14). β strands are labeled a to f, followed by an indication of the domain to which they belong (I or II). The N and C termini are labeled N and C, respectively. Residues of the catalytic dyad, Cys^{144} and His^{41}, are indicated. (**B**) Structure-based sequence alignment of the main proteinases of coronaviruses from all three groups. HCoV, human coronavirus 229E (group I); TGEV, porcine transmissible gastroenteritis virus (group I); MHV, mouse hepatitis virus (group II); BCoV, bovine coronavirus (group II); SARS-CoV, SARS coronavirus (between groups II and III); IBV, avian infectious bronchitis virus (group III). The autocleavage sites of the proteinases are marked by vertical arrows above the sequences. In addition to the sequences of the mature enzymes, four residues each of the viral polyprotein N-terminal to the first and C-terminal to the second autocleavage site are shown. Note the conservation of the cleavage pattern, (small)-Xaa-Leu-Gln ↓ (Ala,Ser,Gly). Thick bars above the sequences indicate α helices (labeled A′ and A to F); horizontal arrows indicate β strands (labeled a to f, followed by the domain to which they belong). Residue numbers for HCoV M^{pro} are given below the sequence; three-digit numbers are centered about the residue labeled. Symbols in the second row below the alignment mark residues involved in dimerization of HCoV and TGEV M^{pro}: open circle (o) indicates only main chain involved; asterisk (*) indicates only side chain involved; plus (+) indicates both main chain and side chain involved. From the almost absolute conservation of side chains involved in dimerization, it can be concluded that SARS-CoV M^{pro} also has the capacity to form dimers. In addition, side chains involved in inhibitor binding in the TGEV M^{pro} complex are indicated by triangles (Δ), and catalytic-site residues Cys^{144} and His^{41} as well as the conserved $Y^{160}MH^{162}$ motif are shaded. (**C**) Cα plot of a monomer of SARS-CoV M^{pro} as model-built on the basis of the crystal structures of HCoV 229E M^{pro} and TGEV M^{pro}. Residues identical in HCoV M^{pro} and SARS-CoV M^{pro} are indicated in red.

HCoV M^{pro} forms a tight dimer in the crystal (the contact interface, which is predominantly between domain II of molecule A and the N-terminal residues of molecule B, is ~1300 Å^2), with the two molecules oriented perpendicular to one another (Fig. 2). Our previous crystal structure of the TGEV M^{pro} (*8*) revealed the same type of dimer. We could show by dynamic light scattering that both HCoV 229E and TGEV M^{pro} exist as a mixture of monomers (~65%) and dimers (~35%) in diluted solutions (1 to 2 mg of proteinase/ml). However, because the architecture of the dimers, including most of the details of intermolecular interaction, is the same in both TGEV M^{pro} (three independent dimers per asymmetric unit) and HCoV 229E M^{pro} (one dimer per asymmetric unit)—that is, in completely different crystalline environments—we believe that dimer formation is of biological relevance in these enzymes. In the M^{pro} dimer, the N-terminal amino acid residues are squeezed in between domains II and III of the parent monomer and domain II of the other monomer, where they make a number of very specific interactions that appear tailor-made to bind this segment with high affinity after autocleavage. This mechanism would immediately enable the catalytic site to act on other cleavage sites in the polyprotein. However, the exact placement of the N terminus also seems to have a structural role for the mature M^{pro}, because deletion of residues 1 to 5 lead to a decrease in activity to 0.3% in the standard peptide-substrate assay (*8*). Nearly all side chains of TGEV M^{pro} and HCoV 229E M^{pro} involved in the formation of this dimer (marked in Fig. 1B) are conserved in the SARS-CoV enzyme, so it is safe to assume a dimerization capacity for the latter as well.

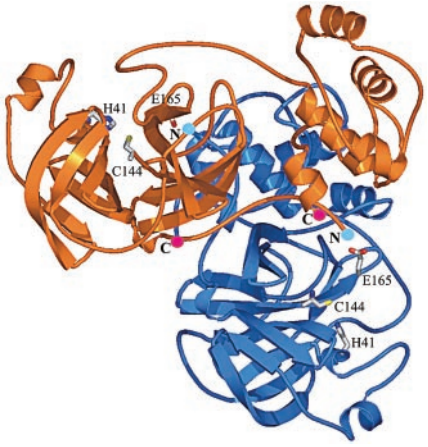


Fig. 2. Dimer of HCoV M^{pro}. The N-terminal residues of each chain squeeze between domains II and III of the parent monomer and domain II of the other monomer. N and C termini are labeled by cyan and magenta spheres and the letters N and C, respectively.

In the active site of HCoV 229E M^{pro}, Cys^{144} and His^{41} form a catalytic dyad. In contrast to serine proteinases and other cysteine proteinases, which have a catalytic triad, there is no third catalytic residue present. HCoV 229E M^{pro} has Val^{84} in the corresponding position (Cys in SARS-CoV M^{pro}), with its side chain pointing away from the active site. A buried water molecule is found in the place that would normally be occupied by the third member of the triad; this water is hydrogen-bonded to His^{41} Nδ1, Gln^{163} Nε2, and Asp^{186} Oδ1 (His, His, and Asp in both SARS-CoV and TGEV M^{pro}).

To allow structure-based design of drugs directed at coronavirus M^{pro}s, we sought to determine the exact binding mode of M^{pro} substrates. To this end, we synthesized the substrate-analog CMK inhibitor Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK and soaked it into crystals of TGEV M^{pro}, because these were of better quality and diffracted to higher resolution than those of HCoV 229E M^{pro}. The sequence of the inhibitor was derived from residues P6 to P1 of the N-terminal autoprocessing site of TGEV M^{pro} [SARS-CoV M^{pro} and HCoV 229E M^{pro} have Thr-Ser-Ala-Val-Leu-Gln and Tyr-Gly-Ser-Thr-Leu-Gln, respectively, at the corresponding positions (Fig. 1B)]. X-ray crystallographic analysis at 2.37 Å resolution (*19*) revealed difference density for all residues [except the benzyloxycarbonyl (Cbz) protective group] of the inhibitor in two molecules (B and F) out of the six TGEV M^{pro} monomers in the asymmetric unit (Fig. 3A). In these monomers, there is a covalent bond between the Sγ atom of Cys^{144} and the methylene group of the CMK.

There are no substantial differences between the structures of the enzyme in the free and in the complexed state. The substrate-analog inhibitor binds in the shallow substrate-binding site at the surface of the proteinase, between domains I and II (Fig. 3A). The residues Val-Asn-Ser-Thr-Leu-Gln occupy, and thereby define, the subsites S6 to S1 of the proteinase. Residues P5 to P3 form an antiparallel β sheet with segment 164 to 167 of the long strand eII on one side, and they also interact with segment 189 to 191 of the loop linking domains II and III on the other (Fig. 3A). The functional importance of this latter interaction is supported by the complete loss of proteolytic activity upon deletion of the loop region in TGEV M^{pro} (*8*).

In coronavirus M^{pro} polyprotein cleavage sites, the P1 position is invariably occupied by Gln. At the very bottom of the M^{pro} S1

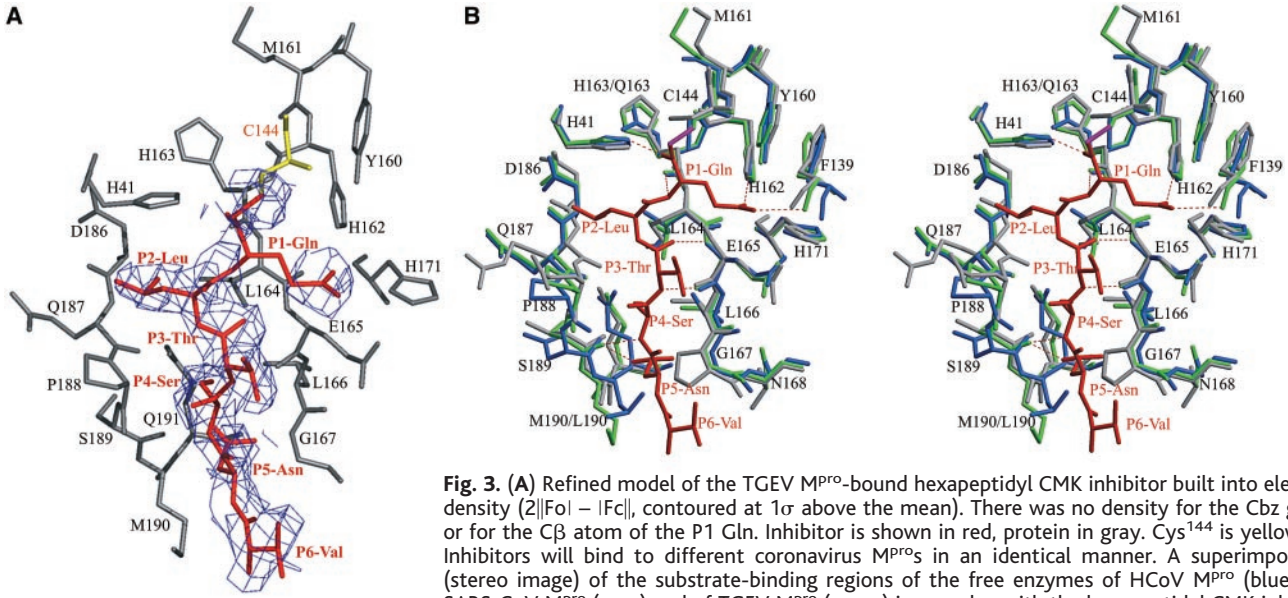


Fig. 3. (**A**) Refined model of the TGEV M^{pro}-bound hexapeptidyl CMK inhibitor built into electron density (2||Fo| – |Fc||, contoured at 1σ above the mean). There was no density for the Cbz group or for the Cβ atom of the P1 Gln. Inhibitor is shown in red, protein in gray. Cys^{144} is yellow. (**B**) Inhibitors will bind to different coronavirus M^{pro}s in an identical manner. A superimposition (stereo image) of the substrate-binding regions of the free enzymes of HCoV M^{pro} (blue) and SARS-CoV M^{pro} (gray) and of TGEV M^{pro} (green) in complex with the hexapeptidyl CMK inhibitor (red) is shown. The covalent bond between the inhibitor and Cys^{144} of TGEV M^{pro} is in purple.

subsite, the imidazole of His^{162} is suitably positioned to interact with the P1 glutamine side chain (Fig. 3, A and B). The required neutral state of His^{162} over a broad pH range appears to be maintained by two important interactions: (i) stacking onto the phenyl ring of Phe^{139} and (ii) accepting a hydrogen bond from the hydroxyl group of the buried Tyr^{160}. In agreement with this structural interpretation, any replacement of His^{162} completely abolishes the proteolytic activity of HCoV 229E and feline coronavirus (FIPV) M^{pro} (*15*, *20*). Furthermore, FIPV M^{pro} Tyr^{160} mutants have their proteolytic activity reduced by a factor of >30 (*20*). All of these residues are conserved in SARS-CoV M^{pro} and, in fact, in all coronavirus main proteinases. Other elements involved in the S1 pocket of the M^{pro} are the main-chain atoms of Ile^{51}, Leu^{164}, Glu^{165}, and His^{171}. In SARS-CoV M^{pro}, Ile^{51} becomes Pro and Leu^{164} is Met, although this is less relevant because these residues contribute to the subsite with their main-chain atoms only (Fig. 3B; side chains involved in specificity sites are marked by "Δ" in Fig. 1B).

Apart from a few exceptions, coronavirus M^{pro} cleavage sites have a Leu residue in the P2 position (*9*). The hydrophobic S2 subsite of the proteinase is formed by the side chains of Leu^{164}, Ile^{51}, Thr^{47}, His^{41}, and Tyr^{53}. The corresponding residues in SARS-CoV M^{pro} are Met, Pro, Asp, His, and Tyr. In addition, residues 186 to 188 line the S2 subsite with some of their main-chain atoms. The Leu side chain of the inhibitor is well accommodated in this pocket. It is noteworthy that SARS-CoV M^{pro} has an alanine residue (Ala^{46}) inserted in the loop between His^{41} and Ile^{51}, but this is easily accommodated in the structural model and does not change the size or chemical properties of the S2 specificity site.

There is no specificity for any particular side chain at the P3 position of coronavirus M^{pro} cleavage sites. This agrees with the P3 side chain of our substrate analog being oriented toward bulk solvent. At the P4 position, there has to be a small amino acid residue such as Ser, Thr, Val, or Pro because of the congested cavity formed by the side chains of Leu^{164}, Leu^{166}, and Gln^{191}, as well as the main-chain atoms of Ser^{189}. These are conserved or conservatively substituted (Leu^{164}→Met^{164}, Ser^{189}→Thr^{189}) in SARS-CoV M^{pro}. The P5 Asn side chain interacts with the main chain at Gly^{167}, Ser^{189}, and Gln^{191} (Pro, Thr, and Gln in the SARS-CoV enzyme), thus involving the loop linking domains II and III; whereas the P6 Val residue is not in contact with the protein. Although the inhibitor used in the present study does not include a P1′ residue, it is easily seen that the common small P1′ residues (Ser, Ala, or Gly) can be easily accommodated in the S1′ subsite of TGEV M^{pro} formed by Leu^{27}, His^{41}, and Thr^{47}, with the latter two residues also being involved in the S2 subsite (Leu, His, and Asp in SARS-CoV M^{pro}). Superimposition of the structures of the TGEV M^{pro}-CMK complex and the free enzyme of HCoV 229E M^{pro} shows that the two substrate-binding sites are basically the same (Fig. 3B). All residues along the P site of the cleft are identical, with the exception of the conservative Met^{190}→Leu^{190} replacement (Ala in SARS-CoV M^{pro}). In other coronavirus species, including the SARS pathogen, M^{pro} residues 167 and 187 to 189 show some substitutions, but because these residues contribute to substrate binding with their main-chain atoms only, the identity of the side chains is less important. Indeed, the substrate-binding site of the SARS-CoV M^{pro} model matches those of its TGEV and HCoV 229E counterparts perfectly (Fig. 3B). Thus, there is no doubt that the CMK inhibitor will bind to the HCoV 229E M^{pro} and SARS-CoV M^{pro}, as well as to all other coronavirus homologs, with similar affinity and in the same way as it does to TGEV M^{pro}.

This proposal as well as the correctness of our structural model for SARS-CoV M^{pro} are strongly supported by cleavage experiments that we carried out with the recombinant SARS virus enzyme (*18*) and the peptide H_2N-VSVNSTLQ↓SGLRKMA-COOH (*21*). This peptide, which represents the N-terminal autoprocessing site of TGEV M^{pro} [the cleavage site is indicated by ↓ (Fig. 1B)] and contains the sequence of our CMK inhibitor, is efficiently cleaved by SARS-CoV M^{pro} but not by its inactive catalytic-site mutant Cys^{145}→Ala^{145} (fig. S1).

Although peptidyl CMK inhibitors themselves are not useful as drugs because of their high reactivity and their sensitivity to cleavage by gastric and enteric proteinases, they are excellent substrate mimetics. With the CMK template structure at hand, we compared the binding mechanism to that seen in the distantly related picornavirus 3C proteinases ($3C^{pro}$s). The latter enzymes have a chymotrypsin-related structure, similar to domains I and II of HCoV 229E M^{pro}, although some of the secondary-structure elements are arranged differently, making structural alignment difficult (sequence identity <10%). Also, they completely lack a counterpart to domain III of coronavirus M^{pro}s. Nevertheless, the substrate specificity of picornavirus $3C^{pro}$s (*22*, *23*) for the P1′, P1, and P4 sites is very similar to that of the coronavirus M^{pro}s (*9*). As shown in Fig. 4, we found similar interactions between inhibitor and enzyme in the case of the human rhinovirus (HRV) serotype 2 $3C^{pro}$ in complex with AG7088 (Scheme 1), an inhibitor carrying a vinylogous ethyl ester instead of a CMK group (*13*). Only parts of the two structures can be spatially superimposed (with a rms deviation of 2.10 Å for 134 pairs of Cα positions out of the ~180 residues in domains I and II). Both

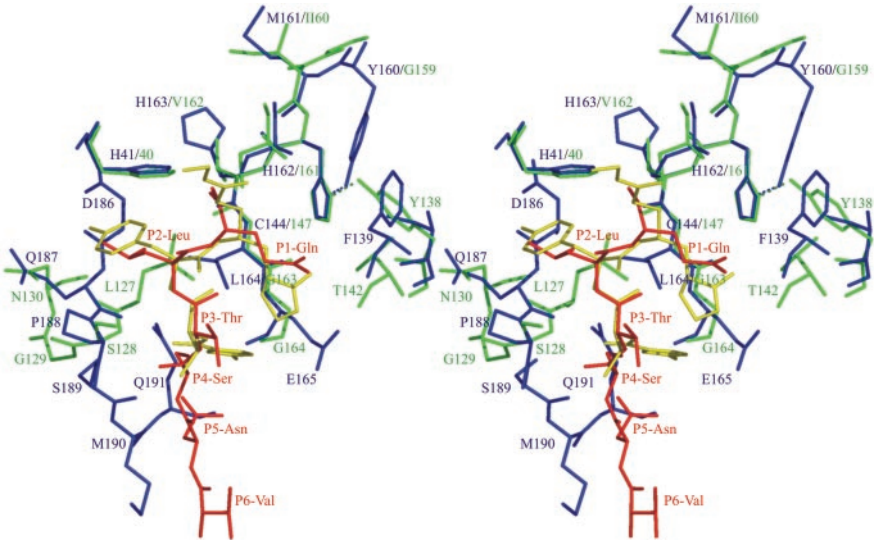
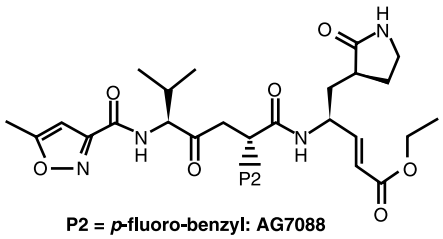


Fig. 4. Derivatives of the antirhinoviral drug AG7088 should inhibit coronavirus M^{pro}s. A superimposition (stereo image) of the substrate-binding regions of TGEV M^{pro} (marine) in complex with the hexapeptidyl CMK inhibitor (red) and HRV2 $3C^{pro}$ (green) in complex with the inhibitor AG7088 (yellow) is shown.



P2 = *p*-fluoro-benzyl: AG7088

Scheme 1.

inhibitors—the hexapeptidyl CMK and AG7088—bind to their respective target proteinases through formation of an antiparallel β sheet with strand eII (Fig. 4). However, completely different segments of the polypeptide chain interact with the substrate analogs on the opposite site: residues 188 to 191 of the loop connecting domains II and III in M^{pro}, as opposed to the short β-strand 126 to 128 in HRV $3C^{pro}$. As a result, the architectures of the S2 subsites are entirely different between the two enzymes; hence, the different specificities for the P2 residues of the substrates (Leu versus Phe). The inhibitor AG7088 has a *p*-fluorophenylalanine side chain (*p*-fluorobenzyl) in this position. Based on molecular modeling, we believe that this side chain might be too long to fit into the S2 pocket of coronavirus M^{pro}, but an unmodified benzyl group would probably fit, as evidenced by Phe occuring in the P2 position of the C-terminal autocleavage site of the SARS coronavirus enzyme (Fig. 1B and table S1). Apart from this difference, the superimposition of the two complexes (Fig. 4) suggests that the P1 and P4 residues of AG7088 (a lactone derivative of glutamine, and 5-methyl-isoxazole-3-carbonyl, respectively) can be easily accommodated by the coronavirus M^{pro}. Thus, AG7088 could well serve as a starting point for modifications that should quickly lead to an efficient and bioavailable inhibitor for coronavirus main proteinases.

The 3D structures for coronavirus main proteinases presented here provide a solid basis for the design of anticoronaviral drugs. The binding modes of substrates and peptidic inhibitors are revealed by the crystal structure of TGEV M^{pro} in complex with the hexapeptidyl CMK. In spite of large differences in binding site architecture of the target enzymes, compound AG7088 binds to human rhinovirus $3C^{pro}$ in much the same orientation as seen for the CMK compound in the binding site of TGEV M^{pro}. This finding indicates that derivatives of AG7088 might be good starting points for the design of anticoronaviral drugs. Because AG7088 has already been clinically tested for treatment of the common cold (targeted at rhinovirus $3C^{pro}$), and because there are no cellular proteinases with which the inhibitors could interfere, prospects for developing broad-spectrum antiviral drugs on the basis of the structures presented here are good. Such drugs can be expected to be active against several viral proteinases exhibiting Gln ↓ (Ser,Ala,Gly) specificity, including the SARS coronavirus enzyme.

References and Notes

1. S. H. Myint, in *The Coronavirdae*, S. G. Siddell, Ed. (Plenum, New York, 1995), p. 389.
2. C. Drosten *et al*., *N. Engl. J. Med.* **348**, 1967 (2003).
3. T. G. Ksiazek *et al*., *N. Engl. J. Med*. **348**, 1953 (2003).
4. N. Lee *et al*., *N. Engl. J. Med.* **348**, 1986 (2003).
5. J. Herold, T. Raabe, B. Schelle-Prinz, S. G. Siddell, *Virology* **195**, 680 (1993).
6. V. Thiel, J. Herold, B. Schelle, S. G. Siddell, *J. Virol*. **75**, 6676 (2001).
7. J. Ziebuhr, J. Herold, S. G. Siddell, *J. Virol*. **69**, 4331 (1995).
8. K. Anand *et al*., *EMBO J*. **21**, 3213 (2002).
9. J. Ziebuhr, E. J. Snijder, A. E. Gorbalenya, *J. Gen. Virol*. **81**, 853 (2000).
10. A. Hegyi, J. Ziebuhr, *J. Gen. Virol.* **83**, 595 (2002).
11. M. Marra *et al*., *Science* **300**, 1399 (2003) (10.1126/science.1085953); see also www.bcgsc.ca/bioinfo/SARS/.
12. P. A. Rota *et al*., *Science* **300**, 1394 (2003) (10.1126/science.1085952).
13. D. A. Matthews *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11000 (1999).
14. The crystallographic *R* factor is 19.8%; R_{free} = 28.0%. For a description of structure elucidation, see the supporting online material.
15. J. Ziebuhr, G. Heusipp, S. G. Siddell, *J. Virol.* **71**, 3992 (1997).
16. The construct of HCoV M^{pro} used in this work lacks two amino acid residues from the C terminus. HCoV $M^{pro}\Delta$(301-302) (with residues 301 and 302 deleted) has the same enzymatic properties as full-length HCoV M^{pro} but yields much superior crystals. In the structure of full-length M^{pro}, residues 301 and 302 are disordered and are not seen in the electron density.
17. S. G. Siddell, in *The Coronavirdae*, S. G. Siddell, Ed. (Plenum, New York, 1995), p. 1.
18. The recombinant SARS-CoV M^{pro} used corresponds to pp1a/pp1ab residues 3241 to 3544 of SARS-CoV strain TOR2 (National Center for Biotechnology Information accession number AY274119).
19. The crystallographic *R* factor is 19.1%; R_{free} = 23.5%. For a description of structure elucidation, see the supporting online material.
20. A. Hegyi, A. Friebe, A. E. Gorbalenya, J. Ziebuhr, *J. Gen. Virol*. **83,** 581 (2002).
21. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
22. H. G. Kräusslich, E. Wimmer, *Annu. Rev. Biochem*. **57**, 701 (1988).
23. M. D. Ryan, M. Flint, *J. Gen. Virol.* **78**, 699 (1997).

24. We thank A. Ulrich for providing peptide synthesis facilities and the staff of ELETTRA (Sincrotrone Trieste, Italy) for help with data collection. Access to this research infrastructure was supported by the European Commission (contract HPRI-CT-1999-00033). The Institute of Molecular Biotechnology Jena/University of Hamburg/European Molecular Biology Laboratory beamline X13 at Deutsches Elelctronersynchrotron is supported by the German Federal Ministry for Education and Research (grant KF1GKA). This work was supported by grants from the Deutsche Forschungsgemeinschaft to R.H. (Hi 611/2) and J.Z. (Zi 618/2). R.H. thanks the Fonds der Chemischen Industrie. Atomic coordinates for the crystal structures of HCoV 229E M^{pro}, the inhibitor complex of TGEV M^{pro}, and the model of SARS-CoV M^{pro} have been deposited with the Research Collaboratory for Structural Bioinformatics Protein Data Bank under accession numbers 1P9S, 1P9U, and 1P9T. 1P9T can be downloaded from ftp://ftp.rcsb.org/pub/pdb/data/structures/models/current/pdb/p9.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1085658/DC1
Materials and Methods
Fig. S1
Tables S1 to S4
References

14 April 2003; accepted 12 May 2003
Published online 13 May 2003;
10.1126/science.1085658
Include this information when citing this paper.